Modern Database Systems
Seminar

Summer Term 2024
Kick-Off
Course Goals

Primary goals:
- Learn how to write a scientific paper
- Learn how to give a scientific talk

Secondary goals:
- Practice to read scientific papers
- Understand modern cloud database system architectures
- Have interesting discussions with your peers
What you get out of this course

In this course you will get:

- At the beginning of the seminar we will give three introduction lectures:
  - Introduction to distributed cloud databases
  - How to write a paper
  - How to give a presentation
- Two peer reviews for your paper
- Presentation feedback from your peers
Prerequisites

Hard prerequisite:
- Introduction to Fundamentals of Databases (IN0008) or equivalent

Beneficial previous knowledge:
- Query Optimization (IN2219)
- Database Systems on Modern CPU Architectures (IN2118)
- Most courses at our chair
Your Deliverables

During the course each of you will create:

- An extended abstract that summarizes your papers topic and focus (~1 page)
- A research paper **draft** (5-8 pages + references)
- 2 peer reviews for your fellow students (~1 page each)
- A research paper (5-8 pages + references)
- A pre-session protocol that shows your preparation for the presentation. (~1 page)
- A presentation (13-16 minutes)
- Meaningful contribution to the discussion in your presentations session
Grading

Rough estimate of grading contributions:

- $\approx 50\%$ Paper
- $\approx 30\%$ Presentation
- $\approx 10\%$ Peer reviews
- $\approx 10\%$ Session preparation protocol and discussion

This is subject to change!
Topics (1)

Preliminary list of topics:

- Early query optimizers for distributed database systems
- Anyblob: Efficient data processing on cloud instances using remote BLOB storage
- Hybrid Client Server Query Processing
- How to do distributed joins fast: NeoJoins and evolution
- Flowjoin: How to handle skew in distributed joins
- Snowflake: A Cloud Unicorn
- Redshift
- Polaris: Save all intermediate results, scale infinitely
- PolarDB
Topics (2)

- Presto / Trino: Metas Distributed Engine
- Velox: Metas Other Distributed Engine
- Lambada: Completely Serverless
- MemSQL Query Optimizer
- Vertica Query Optimizer: Everything is different when you fragment vertically
- Eigen: Alibabas Cluster Scheduling
- Optimizing Data Placement
- Aurora and Scyper: Enabling HTAP with log replication
- Socrates: Scaling OLTP
- PolarDB transactions

Your ideas for related topics are very welcome!
Timeline

Preliminary timeline:

- Mo 15.04.2024 introduction lecture 01
- Mo 22.04.2024 introduction lecture 02 | submit paper preferences
- Mo 29.04.2024 introduction lecture 03
- Mo 06.05.2024 | submit extended abstract
- Mo 13.05.2024
- Mo 20.05.2024 Pfingstferien
- Mo 27.05.2024
- Mo 03.06.2024 | submit paper draft
- Mo 10.06.2024 presentation session 01
- Mo 17.06.2024 presentation session 02 | submit peer reviews
- Mo 24.06.2024 presentation session 03
- Mo 01.07.2024 presentation session 04
- Mo 08.07.2024 presentation session 05
- Mo 15.07.2024 presentation session 06 | submit final paper
Organization

• Attendance to all sessions in presence is mandatory
• You have to write your paper in LaTeX using the VLDB template (typst is okay if you find an indistinguishable template)
• Communication will take place using our Mattermost instance
Matching

Register for the course through the matching platform (https://matching.in.tum.de/)
Send us an application to the email riegerm@in.tum.de using the subject [mdb24] kickoff containing the following:

- Your matriculation number
- Your motivation for the seminar in at most two short sentences
- Relevant courses you took and grades
- A pdf in VLDB template style including only a title, your name as author, and some sample text
- Keep your email short!

Deadline: 14.02.2024
Contact

- Website: http://db.in.tum.de/teaching/ss24/seminarModernDatabaseSystems
- Please send questions to: mds@mailkemper.in.tum.de
- Tobias Götz: goetzt@in.tum.de
- Maximilian Rieger: riegerm@in.tum.de
- Send application to: riegerm@in.tum.de